Biophysical Journal Volume 102 January 2012 333–341

Quantifying the Importance of Protein Conformation on Ligand Migration in Myoglobin

Nuria Plattner[†] and Markus Meuwly[†‡*]
[‡]Department of Chemistry, University of Basel, Basel, Switzerland; and [†]Chemistry Department, Brown University, Providence, Rhode Island

ABSTRACT Myoglobin (Mb) is a model system for ligand binding and migration. The energy barriers (ΔG) for ligand migration in Mb have been studied in the past by experiment and theory and significant differences between different approaches were found. From experiment, it is known that Mb can assume a large number of conformational substates. In this work, these substates are investigated as a possible source of the differences in migration barriers. We show that the initial structure significantly affects the calculated ΔG for a particular transition and that fluctuations in barrier heights $\delta\Delta G$ are of similar magnitude as the free energy barriers themselves. The sensitivity of ΔG to the initial structure is compared to other sources of errors. Different protein structures can affect the calculated ΔG by up to 4 kcal/mol, whereas differences between simple point charge models and more elaborate multipolar charge models for the CO-ligand are smaller by a factor of two. Analysis of the structural changes underlying the large effect of the conformational substate reveals the importance of coupling between protein and ligand motion for migration.

INTRODUCTION

Myoglobin is one of the best characterized proteins, both experimentally and by using various types of simulation techniques, and serves as a model system for studying ligand binding, unbinding, and migration (1). Despite intense work, fundamental physico-chemical properties, such as the CO rebinding and migration barriers after photodissociation, are still poorly understood. Although the different pockets accessible to small diatomic ligands are well characterized by experiment (2–5) and theory/computer simulations (6–9), the pathways between the pockets and the energy barriers associated with them are more debatable. A full characterization of these properties requires direct sampling of the entire free energy surface. A considerable step toward this goal has recently been presented where several trajectories of 90 ns (with eight CO molecules each) have been analyzed to identify ligand entry pathways from the solvent. Despite such a serious effort, no free energy profiles were reported because most transitions between pockets are still rare events and occur only once per trajectory (7).

In principle, any sampling strategy (molecular dynamics (MD) or Monte Carlo (MC) and variants thereof) can be used to characterize the distribution of protein conformations and ligand positions. However, for a full understanding of the migration process it is not only important to consider thermodynamic distributions of states, but also migration kinetics. This is one of the advantages of MD over MC simulations as, in addition to providing efficient sampling from the partition function, they also allow to estimate the timescale on which a system evolves from a given initial state into other states accessible on the timescale of a simulation. Therefore, MD simulations can be used to investigate both kinetic and thermodynamic properties. Although MD simulations allow to estimate timescales in principle, in practice this is difficult for processes occurring on very long timescales. In such cases, a large number of time steps are required, since in MD the propagation time is dictated by the motion of the fastest degrees of freedom, which is on the order of femtoseconds. Ligand (carbon monoxide) migration in Myoglobin (Mb) takes place on a 100 ns timescale, as has been demonstrated in several experimental studies, which makes it almost impossible to directly sample with straightforward MD simulations for a statistically significant number of events (5,10,11). In addition, the migration process does not evolve uniformly in time. It can be divided into a number of individual steps, corresponding to migrations between adjacent pockets. Most of the time the ligand resides in a given pocket, while transitions between pockets occur rapidly but rarely. Therefore, for direct MD sampling, the sampling time is determined by the occurrence of these rare transition events. To circumvent this problem and reduce the sampling time, enhanced sampling methods are often used in combination with simplified interaction potentials (8,12–14). The results obtained by different methods show significant differences in free energy barriers.

Since CO-migration cannot be studied directly by MD simulations on its intrinsic timescale, a number of questions concerning the migration mechanism remain open. One of these questions is the influence of the protein structure on ligand migration. As explained above, the ligand remains in a given pocket for a longer time (typically nanoseconds) whereas transitions between pockets occur on comparatively short timescales (typically femtoseconds). During the time the ligand spends in a given pocket, the protein structure changes continuously from its given initial

Submitted May 4, 2011, and accepted for publication October 26, 2011.
*Correspondence: m.meuwly@unibas.ch
Editor: Gregory A. Voth.

0006-3495/12/01/0333/9 $2.00
doi: 10.1016/j.bpj.2011.10.058

structure into a number of nearly isoenergetic conformational substates. For Mb, conformational substates were first postulated based on the experimentally observed nonexponential time dependence of small molecule binding (15) and later confirmed by other experiments (10,11,16) and theory (17,18). Structural changes between different substates may not be large, but they are functionally relevant (1). One example for such a conformational transition is the difference between the structures of bound MbCO and deoxy Mb. Another example is the concerted motion of the Heme, the Iron, and helices E and F for rebinding (19). Since these conformational changes have been shown to affect ligand binding kinetics, they are also likely to influence ligand migration, which is difficult to observe directly in experiments.

During the time of a transition the protein visits only a small part of the overall available substate space. This implies that the initial substate may affect ligand migration and associated barriers. If migration barriers vary significantly by starting simulations from different substates it is likely that protein and ligand motion are coupled, i.e., transitions between different pockets are only likely if the protein is in a substate that supports such a transition. If, in contrast, the migration barriers are essentially independent of the initial substate, protein and ligand motion are likely to be uncoupled. The aim of this work is not only to characterize the effect of the structure of the substate on an atomistic level but also to quantify it for different known migration barriers. In a complete description of ligand migration the existence and influence of conformational substates has to be taken into account. However, direct simulations would be extremely costly and are still not possible for a statistically significant number of transitions.

Positions visited by the ligand in MC sampling are thermodynamically favorable but may still be very unlikely to occur in practice since pathways connecting them to other ligand positions may not be favorable. For these reasons, MC and MD sampling are used to determine ligand positions in different parts of the substate space that provide initial positions for umbrella sampling from which quantitative data on the free energy barriers can be obtained. The possible positions for unbound ligands inside Mb have been previously characterized by experiment and simulations (2,7–9,12,20) and the pockets and pathways connecting them are shown in Fig. 1.

To compare the order of magnitude of this effect to an alternative and more technical source of differences, different levels of accuracy to represent the ligand electrostatics are evaluated. In a number of recent simulations, electrostatics is either neglected (12) or represented only by a small dipole moment (8,13,14). Here, the consequences of this for the free energy barriers separating neighboring metastable states are investigated by comparing free energy profiles and barriers obtained with different interaction potentials.

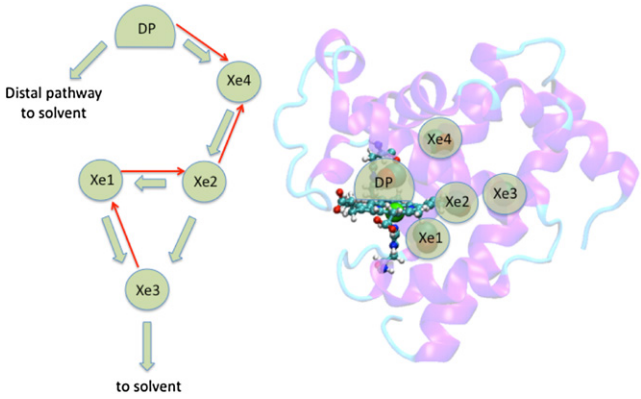

FIGURE 1 Myoglobin structure with pockets and pathways. (*Left*) Pathways between the pockets. Initially the CO molecule is located in the distal pocket (*DP*). (*Right*, *green arrows*) Possible migration pathways from the distal pocket to the solvent; (*red arrows*) reaction coordinates chosen for umbrella sampling.

COMPUTATIONAL METHODS

To obtain free energy profiles for the transitions between the pockets, different simulation methods are combined: umbrella sampling is used to sample the free energy profiles corresponding to the different transitions between the pockets. For this purpose, initial conditions for the CO ligand inside the various pockets are required. These initial conditions are obtained from two different methods: A), Free dynamics, where a spontaneous transition from the initial site in the distal Heme pocket to the Xe4 pocket can be observed after ~1 ns. B), Molecular dynamics combined with Monte Carlo (MC/MD) simulations, where favorable sites include the experimentally characterized Xe1–Xe4 pockets, as well as two pockets (Phantom 1 and Phantom 2, i.e., Ph1 and Ph2) that have been previously found from long (90 ns) MD simulations (8).

Simulation protocols

The computational setup for the molecular dynamics (MD) simulations closely follows that of previous studies of photodissociated MbCO (21). Briefly, hydrogen atoms were added to the x-ray structure (Protein DataBank reference 1MBC) (22), and the heme pocket was fully solvated in a box of dimensions $56 \times 52 \times 40$ Å. Simulations were carried out for the $His_{\varepsilon}64$ protonation state, which is the more likely state (23–27). The solvated structure was equilibrated for 100 ps in the *NPT* ensemble using a time step of 1 fs. For all simulations described below, nonbonded interactions were treated with a 12 Å cutoff and hydrogen atoms were constrained with the SHAKE algorithm (28).

To determine transition barriers between neighboring sites (DP↔Xe4, Xe4↔Xe2, Xe2↔Xe1, and Xe1↔Xe3), umbrella sampling simulations were carried out. Initial structures were obtained following two different simulation protocols, A and B (see the Supporting Material for details),

after which umbrella sampling was carried out according to simulation protocol C). Briefly, these protocols are:

A. "Structure A": Four MD simulations of 1 ns in the *NPT* ensemble starting from the equilibrated structure. Coordinates and velocities at the end of one of the 1-ns simulation with a spontaneous transition to the Xe4 pocket were used.
B. "Structure B": Mixed MC/MD simulations, carried out in a staged fashion. MC moves (translation and rotation) were applied to the CO molecule only, whereas MD was used for the entire system. From structures classified according to the CO position, three structures with the CO in Xe4, and one structure with CO in Xe2 and Xe1, were used, respectively.
C. Umbrella sampling (29): Umbrella sampling was used to sample transitions between pockets (for details, see the Supporting Material). Reaction coordinates used for each transition are summarized in Table 1. As was previously shown, unbiased MD simulations provide a meaningful way for defining reaction coordinates (30–32). The final free energy surfaces were constructed using the data obtained with $k_{umb} = 1$ kcal/mol/Å^2 for all windows if sufficient sampling was obtained and with $k_{umb} = 2$ kcal/mol/Å^2 in the remaining cases. Adjacent free energy segments were joined by the weighted histogram method (33).

Comparison of different interaction potentials

For the CO ligand, four different interaction potentials were used: a conventional two-point charge model (model I) (34,35), a fixed three-point charge model (model II, accurately captures the CO quadrupole moment) (36), and a model using fluctuating atomic multipole moments up to octopole (model III). For the van der Waals (vdW) parameters, standard CHARMM parameters with $\frac{\sigma_C}{2} = 2.1$ Å, $\frac{\sigma_O}{2} = 1.7$ Å, $\varepsilon_C = -0.11$ kcal/mol, and $\varepsilon_C = -0.12$ kcal/mol are compared to radii reduced by 10% (model IV), using model III for the electrostatics. For the CO bond, the spectroscopically accurate RRKR potential was used throughout (37). For comparing the different potentials, structure A was used for models I–III and additional simulations were carried out with structure B using models II and III.

RESULTS

Influence of different initial structures

Barriers for different initial structures are compared for four known transitions: DP↔Xe4, Xe4↔Xe2, Xe2↔Xe1, and Xe1↔Xe3. The free energy profiles are reported in Fig. 2 and numerical values are summarized in Table 2. For the DP↔Xe4 transition, a total of four (one from simulation protocol A and three from B) initial structures have been used and yield four different free energy profiles (see Fig. 2, *upper left*), with barrier heights ranging from 0.8 to 2.8 kcal/mol for DP→Xe4. For transitions Xe4↔Xe2, Xe2↔Xe1, and Xe1↔Xe3, two different structures are compared, one from protocol A, and one from B (see Computational Methods). Differences in the forward and reverse barriers along Xe4↔Xe2 (see Fig. 2) are small compared to the barrier height (0.9 kcal/mol for Xe4→Xe2). On the other hand, for the Xe1↔Xe2 and the Xe3↔Xe1 transitions, differences between 3 and 4 kcal/mol are observed in at least one of the transition directions.

It is interesting to note that two previous simulations for the DP→Xe4 transition in sperm whale and horse Mb, respectively, find 5.4 and 1.3 kcal/mol (13,20). Both simulations were carried out with metadynamics and point charges on the CO molecule and thus it can be expected that a substantial amount of the 4.1 kcal/mol difference originates from the different protein structures used in the simulations. This is also in accord with the results found in this article (see Fig. 2).

Analysis of structural rearrangement

The rearrangements during the umbrella sampling simulations following the different protocols are analyzed to characterize the structural changes. The analysis is carried out based on comparing per-residue root mean-square deviations (RMSDs). After reorienting two structures to their minimum overall RMSD with respect to the Cα atoms, RMSDs are calculated for the heavy atoms of each residue.

TABLE 1 Computational techniques and reaction coordinates for umbrella sampling between different pockets

Pathway	Initial position	Initial structure	Reaction coordinate q
DP → Xe4	Xe4	MD or MC/MD	Fe ↔ CO
Xe4 → Xe2	Xe4	MD or MC/MD	Phe^{138}* ↔ CO
Xe2 → Xe1	Xe2	Xe4 → Xe2 pathway	His^{93†} ↔ CO
Xe1 → Xe3	Xe1	Xe2 → Xe1 pathway	Trp^{14}/Ile^{75}/Leu^{76}/Leu^{135‡} ↔ CO

If the reaction coordinate is defined as the distance between more than two atoms, average distances over the atoms are used, i.e., the reaction coordinate is the distance between the geometric centers of the two atom groups defining it. In the third column, "Xe4→Xe2 pathway" corresponds to a structure taken from the previous sampling of the Xe4→Xe2 transition with the umbrella being placed at the Xe2 minimum; the same applies to "Xe2→Xe1 pathway".
*Carbon atoms only.
†Atom types CG, CE1, CD2, ND1, NE2.
‡Trp^{14} CD2, CE2, CZ2, CH2; Ile^{75} C, CA, CB; Leu^{76} C, Leu^{135} N, CG, CD2.

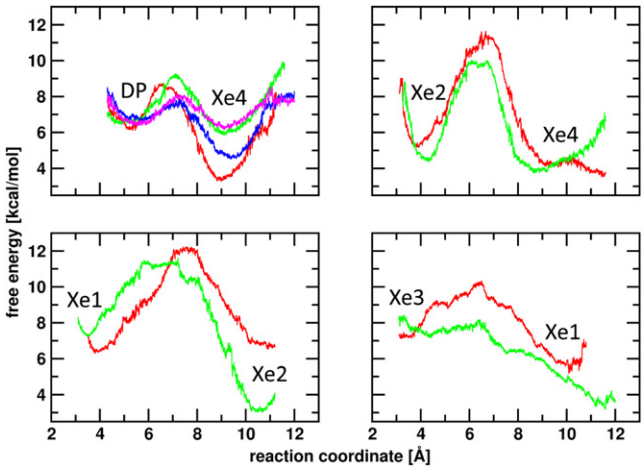

FIGURE 2 Free energy profiles for transitions between pockets from different initial structures using model III for CO. (*Upper left*) Xe4→DP; (*upper right*) Xe4→Xe2; (*lower left*) Xe2→Xe1; (*lower right*) Xe1→Xe3. (*Red* and *green lines*) Different initial structures from simulation protocols A and B (see Computational Methods) and two additional independent initial structures for DP→Xe4 transition (*blue* and *magenta*).

For two coordinate sets a and b involving a specific residue, RMSD (a,b) is defined as

$$RMSD(a,b) = \sqrt{\left(\frac{\sum_{i=1}^{N} r_i^2}{N}\right)},$$

with N the number of heavy atoms in the residue and r_i the distance between the absolute coordinates of atom i in sets a and in b, respectively. The RMSDs are averaged over several structures at the start and the end of a trajectory. For a given residue, the final average $\overline{\text{RMSD}}_{s,e}$ between starting (s) and ending (e) structures is

$$\overline{\text{RMSD}}_{s,e} = \frac{1}{n_{umb}} \sum_{u=1}^{n_{umb}} \frac{\sum_{n=1}^{n_{start}} \sum_{m=1}^{n_{end}} RMSD(n,m)}{n_{start} * n_{end}}, \quad (1)$$

where n_{umb} is the number of umbrella windows along the given reaction coordinate, n_{start} is the number of structures at the start of a trajectory, and n_{end} is the number of structures at the end of the trajectory. In the following, $n_{start} = n_{end} = 3$, corresponding to three structures at the beginning of the equilibration (separated by 1 ps), and three end structures of the same trajectory, also separated by 1 ps. For a given residue, $\overline{\text{RMSD}}_{s,e}$ therefore corresponds to an RMSD average over nine structure pairs and over the entire umbrella simulation for a given transition (i.e., several hundred structures). The results for the Xe4↔Xe2 (*upper panel*) and for the Xe2↔Xe1 (*lower panel*) transition are shown in Fig. 3 (*red* and *green lines*, referring to structures A and B, respectively).

To understand the large effect of the initial structure on the free energy profiles, the same RMSD analysis is used to quantify conformational differences between runs starting from different initial structures A and B. Deviations $\overline{\text{RMSD}}_{s,e}^{(A,B)}$ are evaluated between three starting and final structures of trajectories initiated from structures A and B, respectively. The results are compared to the previous results in Fig. 3 (where RMSDs between structures at the start and at the end of a transition are represented by *blue* and *magenta lines*, respectively). The comparison shows that structural changes within one simulation are significantly smaller than the structural differences between simulations using

TABLE 2 Comparison of activation free energies (in kcal/mol) for transitions between neighboring sites from this work and from the literature

Site	DP → Xe4	Xe4 → Xe2	Xe2 → Xe1	Xe1 → Xe3
Model I and structure A	4.7	8.5	7.9	6.7
Model II and structure A	2.0	7.3	5.5	4.7
Model II and structure B	2.3	7.6	8.4	4.9
Model III and structure A	2.5	6.9	5.2	4.5
Model III and structure B (*i*)	2.8	6.0	8.1	4.4
Model III and structure B (*ii*)	0.8			
Model III and structure B (*iii*)	1.6			
Banushkina and Meuwly (31)*	4.3	—	—	—
Banushkina and Meuwly (32)†	1.3	—	—	—
Ceccarelli et al. (20)‡	5.4	2.9	5.3	5.4
Bossa et al. (8)§	2.4	—	3.0	5.3
Cohen et al. (12)¶	3.5	—	—	—
Nishihara et al. (13)‖	1.3	2.6	1.4	

Structures B (*i*), (*ii*), and (*iii*) correspond to different initial structures obtained with simulation protocol B.
*Umbrella sampling, three-point fluctuating charge model; most comparable to this work.
†Free dynamics; three-point fluctuating charge model; free energy based on two-dimensional diffusion.
‡Metadynamics; point charges.
§Potential of mean force method based on free dynamics; three-point CO model.
¶Implicit ligand sampling; no electrostatics on CO.
‖Metadynamics; point charges; Horse Myoglobin.

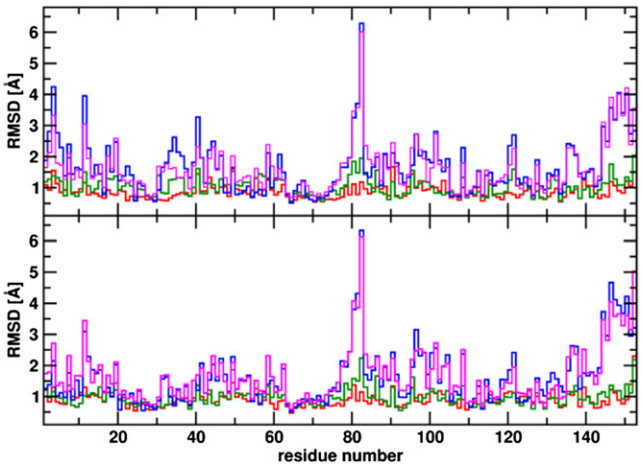

FIGURE 3 RMSD per residue for different averaged structure pairs. (*Red lines*) RMSDs between the starting point and the end-point of a simulation for initial structure A. (*Green lines*) Structure B. (*Blue* and *magenta lines*) Comparison between structures A and B. (*Blue*) Starting points of the trajectories; (*magenta*) end-points of the trajectories. (*Upper panel*) Data for the Xe4→Xe2 transition. (*Lower panel*) Data for the Xe2→Xe1 transition.

different initial structures, despite using identical umbrella potentials. Fig. 4 reports projections of $\overline{\text{RMSD}}_{s,e}^{(A,B)}$ onto the protein structure for the Xe4↔Xe2 transition. Residues with large RMSDs are solvent-exposed whereas residues with low RMSDs are located in less exposed regions and closer to the Heme group.

To evaluate the convergence of free energy profiles in terms of their equilibration under a given umbrella potential, structural changes in umbrella sampling simulations are also analyzed as a function of time. For this purpose, residue

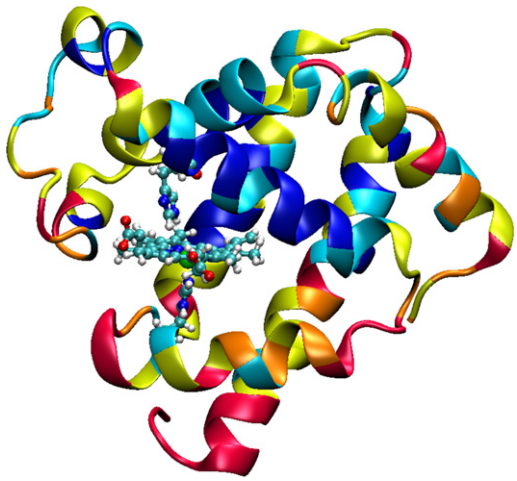

FIGURE 4 Myoglobin structure with color-coded residue RMSDs between structures A and B. The data are evaluated based on the Xe4→Xe2 transition. Color-coding: RMSD < 1 Å (*dark blue*); 1–1.4 Å (*light blue*); 1.4–1.8 Å (*yellow*); 1.8–2.2 Å (*orange*); >2.2 Å (*red*).

displacements $\overline{\text{RMSD}}_{t,t+1}$ for one simulation are calculated between pairs of successive structures, separated by 1 ps, for times t and $t + 1$. The RMSDs are again averaged over the entire reaction coordinate and are shown in Fig. 5 for two groups of residues: one group contains residues within 10 Å of CO, the other group contains residues within 10–15 Å. It is found that the variation of residue RMSDs is largest for the first picosecond, $\overline{\text{RMSD}}_{0,1}$. For most residues, the RMSD between subsequent frames levels off at ~0.4 and 0.7 Å within the first few picoseconds. This corresponds to random fluctuations around an equilibrium position. For some residues, located predominantly in the first residue layer around CO, longer equilibration is required, but most of the relaxation dynamics is complete after 10 ps that corresponds to the time allowed for equilibration during a window in umbrella sampling. Therefore, the free energy profiles can be considered converged with respect to structural rearrangements under a given umbrella potential.

Structural origin of varying migration barriers

For characterizing structural origins in free energy barrier variations, residue space is formally divided into two parts: a (spatially) local region around the transition pathway formed by on-pathway residues that are in vdW contact with the ligand during its passage from one local minimum

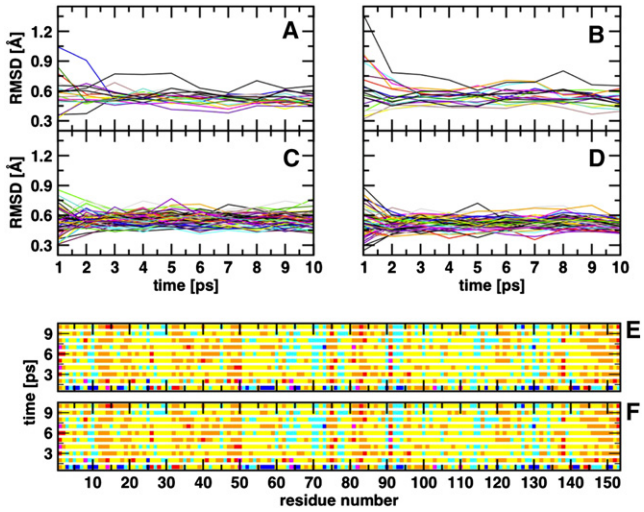

FIGURE 5 RMSD per residue between subsequent structures during the 10-ps umbrella sampling equilibration separated by 1 ps. Panels *A–D* show selected residues, panels *E* and *F* show an overview over all residues. Selected residue part: panels *A* and *B* show all residues within 10 Å of CO for the given transition, panels *C* and *D* show all residues between 10 and 15 Å of CO. Panels *B* and *D* show the data for the Xe4→Xe2 transition for structure B; panels *A* and *C* show the data for the Xe2→Xe1 transition for structure A. Residue overview part: the colors indicate the range of RMSD for a given residue at a given time: RMSD < 0.4 Å (*dark blue*); 0.4–0.5 Å (*light blue*); 0.5–0.6 Å (*yellow*); 0.6–0.7 Å (*orange*); 0.7–0.8 Å (*red*); >0.8 Å (*magenta*). Panel *E* shows the data for the Xe4→Xe2 transition for structure B; panel *F* shows the data for the Xe2→Xe1 transition for structure A.

to the next, and a global region containing all other residues of the protein. Off-pathway residues are residues beyond the vdW contact but important in their ability to respond to changes in the on-pathway residues. The contribution of on-/off-pathway residues is quantified for the Xe4↔Xe2 and the Xe2↔Xe1 transition from umbrella sampling simulations with harmonic mass constraints (10 kcal/mol/Å^2) applied to on-pathway residues (all C_α atoms within 7 Å of the CO molecule) and to off-pathway residues (all other C_α atoms).

Free energy profiles from simulations without and with constraints are reported in Fig. 6, *A* and *B*. For this evaluation, structure A was used. Therefore, the unconstrained free energy profiles are identical to those shown (in *red*) in Fig. 2. Overall, both on- and off-pathway residues significantly affect the forward and reverse barrier heights, $\Delta G_{Xe2\leftrightarrow Xe4}$ and $\Delta G_{Xe1\leftrightarrow Xe2}$. Differences between barrier heights from unconstrained and constrained simulations range from 1.5 kcal/mol (Xe2↔Xe4) to 4 kcal/mol (Xe1↔Xe2). The larger overall differences for the Xe1↔Xe2 transition compared to the Xe2↔Xe4 transition are consistent with the sensitivity of the barriers for simulations with structures A and B, respectively (see Table 2). The free energy changes indicate that rearrangements within the pockets and neighboring residues are possible, which is also in line with the experimentally observed rearrangement of the pockets upon ligand binding (38). This evaluation shows that both, local and global adaptations affect the resulting free energy profile. Constraining C_α atoms from off-pathway residues allows adaptations of side- and main-chain atoms of on-pathway residues. This leads to similar forward and reverse barriers for the Xe2↔Xe4 transition and to a similar forward barrier in the case of the Xe1↔Xe2 transition as in unconstrained simulations. Conversely, constraints on the on-pathway C_α atoms primarily allows the side chains of on-pathway residues to adapt and leads to small forward barriers compared to unconstrained simulations. However, the reverse barriers are equal or larger than in unconstrained ones and correspond to those from off-pathway constrained simulations. Therefore both on- and off-pathway residues affect ligand migration barriers. The degree to which individual regions of the protein influence the transition barriers depends on the particular transition studied. Additional simulations with both types of constraints applied simultaneously show that the effect of on- and off-pathway residues is not additive.

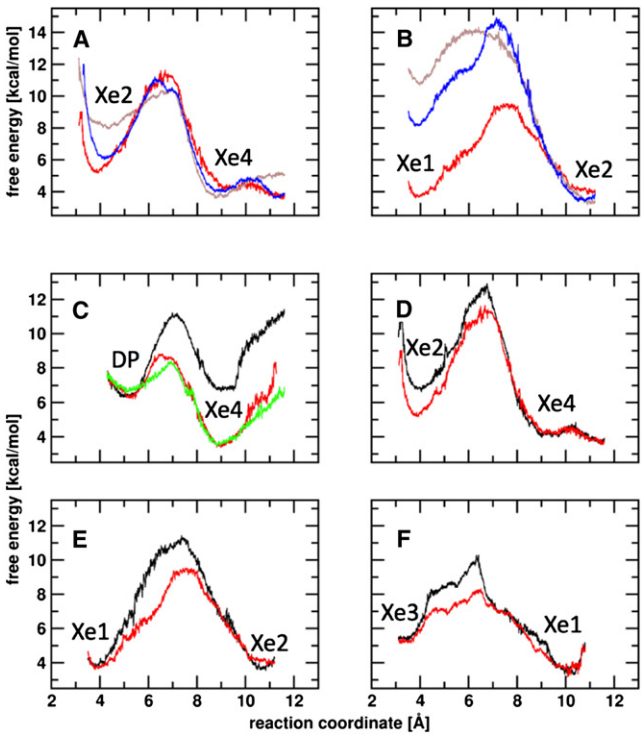

FIGURE 6 Effect of on- and off-pathway constraints on the free energy profiles for the Xe2→Xe4 (*A*) and the Xe1→Xe2 transition (*B*). Results from unconstrained simulations are shown (*red*), free energy curves with on-pathway constraints, on Cα atoms within 7 Å of the CO position, are reported (*brown*), and curves (*blue*) refer to the effect of constraints on off-pathway Cα atoms. Free energy profiles for transitions between pockets from electrostatic models I–IV for CO. Xe4→DP (*C*); Xe4→Xe2 (*D*); Xe2→Xe1 (*E*); Xe1→Xe3 (*F*). Color-codes are for different CO models: model I (*black*, point charges with standard vdW radii), model III (*red*, multipole moments with standard vdW radii), Model IV (*green*, multipole moments with vdW radii reduced by 10%).

Influence of different interaction potentials

To put the magnitude of the above effects into perspective, transition barriers ΔG were calculated from simulations with electrostatic interaction potentials I–III (see Computational Methods) for all four transitions. The results are summarized in Fig. 6 and compared to previous studies (8,12,13,20,31,32) in Table 2. For the simple point charge model (model I), the transition barriers are systematically larger than for the more elaborate multipole model (model III). In some cases the difference is small compared to the barrier height, e.g., for the Xe4↔Xe2 transition where it is ≈0.9 kcal/mol in both directions. In other cases it is more substantial (Xe2↔Xe1, and Xe1↔Xe3, with 2.2 and 2.7 kcal/mol, respectively). Because the difference is systematic, the effect cannot be neglected and is caused by the simplified point charge representation. However, the effect is similar or typically smaller than the effect of different initial structures. Changing the vdW radii of the C and O atoms by 10% does not appreciably affect the transition free energy barrier. Results for the three-point charge model (model II) show only minor differences compared to model III. These findings suggest that a representation of the molecular moments up to quadrupole is required to obtain accurate free energies. This can be achieved by using either a three-point charge model (21,36,39) or atomic multipole moments (40). In contrast, the choice of vdW parameters and molecular moments higher than quadrupole does not

appear to be crucial for free energies. This differs from findings for infrared spectra, where molecular moments up to octopole are important to obtain accurate results (40,41).

DISCUSSION AND CONCLUSIONS

The essential insights gained by this work are summarized below, and further discussed in the following:

1. The initial substate from which a transition is initiated has a substantial effect on ΔG and for the fluctuations, we find $\delta\Delta G \approx \Delta G$. This effect is genuine and typically larger than purely technical differences (e.g., point charge models) in the simulations.
2. For a given ligand position to be favorable, structural rearrangements distributed over the entire protein contribute. This also means, in turn, that the ligand position affects the energy and population of substates. Hence, ligand migration is not passive diffusion but coupled motion between protein and ligand and the coupling strength depends on the position along the reaction coordinate.

The initial conformation considerably affects individual ligand migration barriers (between 2 and 4 kcal/mol). This change in barrier height would amount to 2–3 orders of magnitude in the corresponding transition rate. Therefore it is evident that the transition probability depends on the protein substate from which the transition takes place. One might argue that umbrella sampling leads to locally strained conformations that artificially raise barriers for ligand transfer. However, the analysis of residue adaptation as a function of time (see above) shows that equilibrium is restored on a timescale of a few picoseconds. Nevertheless, some of the sampled conformations are likely to not be part of the transition state ensemble because, due to the combined use of MC and MD methods, it is possible for the ligand to surmount barriers and protein and ligand motions are partially uncoupled. Therefore, conformations that are kinetically inaccessible in a given substate can be sampled. This makes it possible to detect the effect of protein-ligand coupling on the free energy barriers.

Computed barriers depend more on structural aspects than on details of the interaction potentials. The comparison of the different interaction potentials shows a systematic change in ΔG, whereby the barriers obtained using the point charge model are always higher by up to $\approx$2 kcal/mol compared to simulations with a multipolar model. This underlines the need for accurate interaction potentials to calculate reliable free energy barriers. The fact that the influence of the electrostatic model used is typically smaller than the effect of the conformational substate from which simulations are carried out, supports the finding that adequate sampling of initial conditions is necessary in order to obtain realistic free energy profiles.

The sensitivity of ΔG to the initial substate is a consequence of the different timescales on which ligand transitions and the overall protein conformational changes take place. The time for ligand transition between two pockets and the time required to sample a transition using umbrella sampling is significantly shorter than the timescale for overall structural rearrangement and relaxation, i.e., separation of timescales. Analysis of the structural changes shows that rearrangements along one umbrella sampling simulation are significantly smaller than structural differences between different initial structures. Thus, for migration paths with a high barrier, no transitions to an overall protein structure favoring a low-energy passage take place. The protein is indeed frozen in a substate and the ligand has to find alternative means to progress in configurational space. Consequently, each of the conformational substates has its own barrier associated with it. Considering an ensemble N of initial structures i, a simplified picture would associate the average barrier $\langle\Delta G(s)\rangle = \sum_{i=1}^{N}\Delta G^{(i)}(s)$ along a reaction coordinate s with the experimentally determined one, whereas the fluctuation $\delta\langle\Delta G(s)\rangle$ around the average is a measure for the influence of and coupling to, degrees of freedom orthogonal to s. The findings reported in this work for CO migration in Mb show that $\langle\Delta G(s)\rangle$ and $\delta\langle\Delta G(s)\rangle$ are of comparable magnitude and cannot be easily separated.

The analysis of the structural origin of varying migration barriers shows that for a given ligand position to be favorable, structural rearrangements involving more distant residues are required. One of the effects of the biasing umbrella potential is to artificially shorten the timescale for the structural rearrangement necessary to allow the transition to take place and to overcome a barrier along a reaction coordinate s more rapidly. The results presented here support the notion that the modulation by and coupling to degrees of freedom orthogonal to s affects the barriers. Depending on the detailed nature of the couplings and the bias, the timescale shortening may affect residues within different distances of the transition, but the details of the residue adaptation are substate-dependent, i.e., the structural rearrangement is not of the same order of magnitude as structural changes in the protein occurring on a longer timescale. As a consequence, the adapted structure is still relatively similar to the initial structure. Because residues all over the protein affect the free energy of a given ligand position, it also follows that the position of the ligand within the protein can affect the probability of a given substate. However, the change between substates takes place on a longer timescale than the transition between adjacent pockets because it involves motion of the entire protein. Thus, the substate populations are not equilibrated with respect to a given ligand position during the transition time, and typically multiple samplings will be required to converge free energy profiles.

The apparent sensitivity of computed barriers on the geometrical structure of the environment is related to fundamental questions about the mechanism underlying ligand

transport in proteins. Is ligand transport passive or active, i.e., diffusive or coupled (slaved) to the protein motion? If ligand migration was purely diffusive, only an insignificant dependence of the barrier on the initial conformation from a thermalized ensemble would be expected. This is manifestly not the case, despite the fact that the simulation techniques employed allow structural changes of the protein to conformations that would facilitate ligand passage. The importance of fluctuations among conformational substates for the passage of ligands through the protein matrix has been suggested before based on experimental observations and simulations (7,11). However, it should be noted that earlier interpretations primarily focused on the protein providing a rough energy landscape—through the existence of many almost equivalent conformational substates—on which the ligand diffuses, whereas our simulations treat ligand and protein on an equal footing and allow dynamical coupling between protein and ligand dynamics.

In this work the distributed barriers for ligand migration observed from experiments have, for the first time (to our knowledge), been quantified from atomistically detailed simulations. A biologically relevant and functionally important insight from this study is that fluctuations in barrier heights from different simulations can be as large as the barrier itself. This is in agreement to the width of the barrier distributions determined experimentally based on ligand rebinding kinetics. For CO rebinding in native Myoglobin, a barrier range from 6 to 9 kJ/mol has been found for barrier heights of 9–18 kJ/mol (11). For the Myoglobin mutant L29W, activation enthalpies of 6–22 kJ/mol with widths ranging from 2 to 3 kJ/mol were found (16). Because the latter results have been obtained by kinetic studies at different temperatures, a correction for the temperature change would be required for direct comparison. However, the uncorrected results show already that a substantial variation of energy barriers exists. The importance of the overall protein conformation for reaction barriers has been found before in studies of enzyme reactions. In an MD study of proton transfer in triosephosphate isomerase it has been found that the height of the barrier is determined by the configuration of the intramolecular subsystem. The dynamics of this subsystem is coupled to the low frequency fluctuations of the enzyme and, therefore, to its overall configuration at the time of the reaction (42).

In conclusion, our simulations and analysis establish that the structure of a given substate significantly affects the passage of the ligand between two pockets. Analysis of the residue adaption and in particular of the effect of distant residue adaptation shows that the ligand position within the protein affects residues throughout the entire protein and can therefore also affect the population of different substates. As a consequence, the interplay between ligand and protein seems to be characterized by effects involving multiple timescales and correlations between ligand position and protein substate populations. The effect of the ligand on the protein substate ensemble can potentially be either favorable or unfavorable for a given transition. In both cases, it affects ligand migration and therefore, the function of the protein.

SUPPORTING MATERIAL

Simulation Protocols: Generation of Initial Structures is available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(11)05351-3.

The authors gratefully acknowledge financial support from the Swiss National Science Foundation through grant 200021-117810.

REFERENCES

1. Frauenfelder, H., B. H. McMahon, and P. W. Fenimore. 2003. Myoglobin: the hydrogen atom of biology and a paradigm of complexity. *Proc. Natl. Acad. Sci. USA.* 100:8615–8617.
2. Tilton, Jr., R. F., I. D. Kuntz, Jr., and G. A. Petsko. 1984. Cavities in proteins: structure of a metmyoglobin-xenon complex solved to 1.9 Å. *Biochemistry.* 23:2849–2857.
3. Olson, J. S., and G. N. Phillips, Jr. 1996. Kinetic pathways and barriers for ligand binding to myoglobin. *J. Biol. Chem.* 271:17593–17596.
4. Scott, E. E., Q. H. Gibson, and J. S. Olson. 2001. Mapping the pathways for O_2 entry into and exit from myoglobin. *J. Biol. Chem.* 276:5177–5188.
5. Schotte, F., M. Lim, …, P. A. Anfinrud. 2003. Watching a protein as it functions with 150-ps time-resolved x-ray crystallography. *Science.* 300:1944–1947.
6. Elber, R., and M. Karplus. 1990. Enhanced sampling in molecular dynamics: use of the time-dependent Hartree approximation for a simulation of carbon monoxide diffusion through myoglobin. *J. Am. Chem. Soc.* 112:9161–9175.
7. Ruscio, J. Z., D. Kumar, …, A. V. Onufriev. 2008. Atomic level computational identification of ligand migration pathways between solvent and binding site in myoglobin. *Proc. Natl. Acad. Sci. USA.* 105:9204–9209.
8. Bossa, C., M. Anselmi, …, A. Di Nola. 2004. Extended molecular dynamics simulation of the carbon monoxide migration in sperm whale myoglobin. *Biophys. J.* 86:3855–3862.
9. Plattner, N., J. D. Doll, and M. Meuwly. 2010. Spatial averaging for small molecule diffusion in condensed phase environments. *J. Chem. Phys.* 133:044506.
10. Frauenfelder, H., F. Parak, and R. D. Young. 1988. Conformational substates in proteins. *Annu. Rev. Biophys. Biophys. Chem.* 17:451–479.
11. Steinbach, P. J., A. Ansari, …, D. C. Lamb…, 1991. Ligand binding to heme proteins: connection between dynamics and function. *Biochemistry.* 30:3988–4001.
12. Cohen, J., A. Arkhipov, …, K. Schulten. 2006. Imaging the migration pathways for O_2, CO, NO, and Xe inside myoglobin. *Biophys. J.* 91:1844–1857.
13. Nishihara, Y., S. Hayashi, and S. Kato. 2008. A search for ligand diffusion pathway in myoglobin using a metadynamics simulation. *Chem. Phys. Lett.* 464:220–225.
14. Maragliano, L., G. Cottone, …, E. Vanden-Eijnden. 2010. Mapping the network of pathways of CO diffusion in myoglobin. *J. Am. Chem. Soc.* 132:1010–1017.
15. Austin, R. H., K. Beeson, …, V. C. Marshall. 1974. Activation energy spectrum of a biomolecule: photodissociation of carbonmonoxy myoglobin at low temperatures. *Phys. Rev. Lett.* 32:403–405.
16. Ostermann, A., R. Waschipky, …, G. U. Nienhaus. 2000. Ligand binding and conformational motions in myoglobin. *Nature.* 404:205–208.

17. Stein, D. L. 1985. A model of protein conformational substates. *Proc. Natl. Acad. Sci. USA*. 82:3670–3672.

18. Elber, R., and M. Karplus. 1987. Multiple conformational states of proteins: a molecular dynamics analysis of myoglobin. *Science.* 235:318–321.

19. Kachalova, G. S., A. N. Popov, and H. D. Bartunik. 1999. A steric mechanism for inhibition of CO binding to heme proteins. *Science.* 284:473–476.

20. Ceccarelli, M., R. Anedda, …, P. Ruggerone. 2008. CO escape from myoglobin with metadynamics simulations. *Proteins.* 71:1231–1236.

21. Nutt, D. R., and M. Meuwly. 2003. Theoretical investigation of infrared spectra and pocket dynamics of photodissociated carbonmonoxy myoglobin. *Biophys. J.* 85:3612–3623.

22. Kuriyan, J., S. Wilz, …, G. A. Petsko. 1986. X-ray structure and refinement of carbon-monoxy (Fe II)-myoglobin at 1.5 Å resolution. *J. Mol. Biol.* 192:133–154.

23. Johnson, J. B., D. C. Lamb, …, R. D. Young. 1996. Ligand binding to heme proteins. VI. Interconversion of taxonomic substates in carbonmonoxymyoglobin. *Biophys. J.* 71:1563–1573.

24. Rovira, C., B. Schulze, …, M. Parrinello. 2001. Influence of the heme pocket conformation on the structure and vibrations of the Fe-CO bond in myoglobin: a QM/MM density functional study. *Biophys. J.* 81:435–445.

25. Merchant, K. A., W. G. Noid, …, M. D. Fayer. 2003. Structural assignments and dynamics of the A substates of MbCO: spectrally resolved vibrational echo experiments and molecular dynamics simulations. *J. Phys. Chem. B.* 107:4–7.

26. Meuwly, M. 2006. On the influence of the local environment on the CO stretching frequencies in native myoglobin: assignment of the B-states in MbCO. *ChemPhysChem.* 7:2061–2063.

27. Nienhaus, K., J. S. Olson, …, G. U. Nienhaus. 2005. The origin of stark splitting in the initial photoproduct state of MbCO. *J. Am. Chem. Soc.* 127:40–41.

28. Van Gunsteren, W., and H. Berendsen. 1977. Algorithms for macromolecular dynamics and constraint dynamics. *Mol. Phys.* 34:1311–1327.

29. Kottalam, J., and D. A. Case. 1988. Dynamics of ligand escape from the heme pocket of myoglobin. *J. Am. Chem. Soc.* 110:7690–7697.

30. Nutt, D. R., and M. Meuwly. 2004. Ligand dynamics in myoglobin: calculation of infrared spectra for photodissociated NO. *ChemPhysChem.* 5:1710–1718.

31. Banushkina, P., and M. Meuwly. 2005. Free-energy barriers in MbCO rebinding. *J. Phys. Chem. B.* 109:16911–16917.

32. Banushkina, P., and M. Meuwly. 2007. Diffusive dynamics on multidimensional rough free energy surfaces. *J. Chem. Phys.* 127:135101.

33. Kumar, S., D. Bouzida, …, J. M. Rosenberg. 1992. The weighted histogram analysis method for free-energy calculations on biomolecules. *J. Comput. Chem.* 13:1011–1021.

34. Brooks, B. R., R. E. Bruccoleri, …, M. Karplus. 1983. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4:187–217.

35. MacKerell, Jr., A. D., D. Bashford, …, M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 102:3586–3616.

36. Straub, J. E., and M. Karplus. 1991. Molecular dynamics study of the photodissociation of carbon monoxide from myoglobin: ligand dynamics in the first 10 ps. *Chem. Phys.* 158:221–248.

37. Huffaker, J. N. 1976. Diatomic molecules as perturbed Morse oscillators. I. Energy levels. *J. Chem. Phys.* 64:3175–3181.

38. Tomita, A., T. Sato, …, S. Adachi. 2009. Visualizing breathing motion of internal cavities in concert with ligand migration in myoglobin. *Proc. Natl. Acad. Sci. USA.* 106:2612–2616.

39. Nutt, D. R., and M. Meuwly. 2004. CO migration in native and mutant myoglobin: atomistic simulations for the understanding of protein function. *Proc. Natl. Acad. Sci. USA.* 101:5998–6002.

40. Plattner, N., and M. Meuwly. 2008. The role of higher CO-multipole moments in understanding the dynamics of photodissociated carbonmonoxide in myoglobin. *Biophys. J.* 94:2505–2515.

41. Lee, M. W., and M. Meuwly. 2011. On the role of nonbonded interactions in vibrational energy relaxation of cyanide in water. *J. Phys. Chem. A.* 115:5053–5061.

42. Neria, E., and M. Karplus. 1997. Molecular dynamics of an enzyme reaction: proton transfer in TIM. *Chem. Phys. Lett.* 267:23–30.